

Pemetaan Penyebaran Guru di Provinsi Banten dengan Menggunakan Metode *Spatial Clustering* K-Means (Studi kasus : Wilayah Provinsi Banten)

Yohanes Aji Priambodo¹⁾, Sri Yulianto Joko Prasetyo^{*2)}

^{1,2)}Fakultas Teknologi Informasi, Program Studi Teknik Informatika Universitas Kristen Satya Wacana Salatiga

Tingkat pendidikan di Indonesia masih tergolong rendah, salah satu penyumbang peran terbesar dari pendidikan adalah peran guru. Pemerataan guru di Indonesia tidak merata di semua wilayah, sebagian besar guru hanya menempati wilayah perkotaan. Berdasarkan data Badan Pusat Statistik tahun 2013, Provinsi Banten memiliki Indeks Pembangunan Manusia terendah di pulau Jawa. Oleh karena itu penelitian ini bertujuan untuk memetakan penyebaran guru di wilayah Provinsi Banten. Pada penelitian ini pengelompokan menggunakan algoritma K-Means berdasarkan jumlah guru, jumlah murid dan jumlah sekolah pada jenjang pendidikan Sekolah Dasar, Sekolah Menengah Pertama dan Sekolah Menengah Atas. Hasil pengelompokan kemudian dipetakan dengan menggunakan Sistem Informasi Geografis. Hasil dari penelitian ini adalah pemetaan wilayah Provinsi Banten berdasarkan jenjang pendidikan yang memiliki kekurangan, kecukupan dan kelebihan guru sesuai dengan kabupaten/kota. Manfaat dari hasil penelitian ini adalah dapat menjadi saran bagi Dinas Pendidikan Provinsi Banten dalam hal pemerataan guru.

Keywords: Penyebaran Guru, *Clustering*, *Data Mining*, Algoritma K-Means, Sistem Informasi Geografis.

I. PENDAHULUAN

Menurut data dari UNESCO pendidikan di Indonesia menempati peringkat ke-10 dari 14 negara berkembang [1]. Penyumbang terbesar rendahnya mutu pendidikan tentu dari wilayah pedesaan. Ini terjadi karena kurangnya tenaga pendidik handal yang mau tinggal di pelosok-pelosok daerah [2]. Distribusi guru masih belum merata, di samping kualitasnya yang masih rendah. Sebanyak 21% sekolah di perkotaan, 37% sekolah di pedesaan, dan 66% sekolah di daerah terpencil masih kekurangan guru [3]. Guru memiliki peran penting dalam meningkatkan pendidikan karena guru merupakan pelaksana terdepan pendidikan anak-anak di sekolah. Berdasarkan data yang bersumber dari Roadmap Manajemen ASN dan Perencanaan Formasi Tahun 2014, penyebaran guru di tiap daerah di Indonesia tidaklah merata [4].

Pada penelitian ini dilakukan uji pemetaan penyebaran guru di Provinsi Banten. Provinsi Banten dipilih sebagai lokasi penelitian dengan alasan bahwa Provinsi Banten memiliki Indeks Pembangunan Manusia (IPM) terendah di pulau Jawa yaitu sebesar 71,90 berdasarkan data dari Badan Pusat Statistik (BPS) tahun 2013 [5]. Selain itu pada tahun 2016 ini ratusan guru yang mengajar di sejumlah sekolah di Provinsi Banten mulai dari tingkatan SD, SMP, hingga SMA/SMK sudah memasuki masa pensiun sejak Januari hingga Desember tahun ini. Sebanyak 60 persen dari total 88.000 guru yang berstatus Pegawai Negeri Sipil (PNS) berada di wilayah perkotaan di daerah Kabupaten / Kota di Banten [6] Oleh karena itu pemerataan guru perlu dilakukan di Provinsi Banten karena hampir semua kabupaten / kota di

Provinsi Banten mengalami ketidakmerataan dalam penyebaran guru.

Salah satu model yang dapat dilakukan untuk melakukan pemetaan adalah dengan sistem informasi geografis yang dibantu dengan metode *clustering* K-Means untuk pengelompokan dan input data pada sistem informasi geografis. K-Means merupakan salah satu metode data *clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster* / kelompok. Metode ini mempartisi ke dalam *cluster* / kelompok sehingga data yang memiliki karakteristik yang sama (*High intra class similarity*) dikelompokkan ke dalam satu *cluster* yang sama dan yang memiliki karakteristik yang berbeda (*Low inter class similarity*) dikelompokkan pada kelompok yang lain [7]. Dengan menggunakan metode K-Means diharapkan bahwa data dapat diolah sehingga dapat membedakan bagian wilayah yang memiliki kekurangan, kecukupan dan kekurangan guru kemudian dipetakan dengan menggunakan sistem informasi geografis. Sistem Informasi Geografis (SIG) adalah sistem informasi yang berdasar pada data keruangan dan merepresentasikan objek di bumi [8].

Berdasarkan latar belakang masalah, maka akan dilakukan penelitian yang membahas tentang pemetaan penyebaran guru di provinsi Banten dengan menggunakan metode *clustering* K-Means.

II. TINJAUAN PUSTAKA

Penelitian berjudul “Sistem Informasi Geografis Sekolah Menengah Atas (SMA/SMK) di Yogyakarta” membahas tentang Pemetaan SMA/SMK di Yogyakarta yang juga menggunakan Sistem Informasi Geografis. Penelitian ini berfokus pada bagaimana membangun aplikasi SIG yang

dapat memetakan letak SMA atau SMK negeri dan swasta di kota Yogyakarta. Penelitian ini telah menghasilkan aplikasi yang cukup baik, namun memiliki kekurangan yaitu masih berbasis desktop, sehingga informasi yang dihasilkan hanya untuk kalangan internal saja. Kelebihan penelitian ini adalah permasalahan pemetaan SMA/SMK dapat diselesaikan dengan baik menggunakan Sistem Informasi Geografis. Alasan penulis menggunakan Sistem Informasi Geografis karena dapat menyelesaikan masalah yang ada yaitu mengenai pemetaan SMA/SMK di Yogyakarta [9].

Penelitian berjudul “Sistem Informasi Geografis Pemetaan Potensi SMA/SMK Berbasis Web” bertujuan untuk memetakan potensi SMA/SMK di Kabupaten Kebumen yang dapat memberikan informasi tentang SMA/SMK. Hasil dari penelitian ini adalah sebuah *web* yang mampu memetakan potensi SMA/SMK di Kabupaten Kebumen. Kelebihan penelitian ini adalah menggunakan bahasa pemrograman PHP dengan *Framework Codeigniter 2.0.2* serta tampilan *service* dari *Google Maps* sehingga mempermudah pengguna dalam mencari lokasi SMA/SMK. Kekurangan dari penelitian ini adalah belum menggunakan Android yang lebih efisien dan mudah diakses bagi penggunaannya. Sistem Informasi Geografis dengan menggunakan tampilan *service* dari *Google Maps* dipilih karena mampu menampilkan peta tentang lokasi SMA/SMK di Kabupaten Kebumen beserta dengan informasinya [10].

Penelitian berjudul “*Application of K-Means Clustering algorithm for prediction of Students' Academic Performance*” bertujuan untuk memantau prestasi akademik mahasiswa dengan cara mengatur skor sesuai dengan tingkat kinerja mahasiswa. Hasil dari penelitian ini adalah bahwa algoritma K-Means dapat memantau perkembangan mahasiswa di institusi tinggi dan memantau kinerja mahasiswa dari setiap semester untuk memperbaiki hasil akademik di masa yang akan datang. Kelebihan dari penelitian ini adalah dengan metode *clustering* K-Means dapat menyelesaikan masalah yang ada. Penelitian ini belum menggunakan GUI (*Graphical user interface*) untuk menampilkan hasil sehingga menjadi kekurangan saat pengguna ingin mengakses data hasil dari penelitian. Penelitian ini dilakukan dengan menggunakan metode *clustering* K-Means karena metode ini cocok dalam menyelesaikan permasalahan untuk memantau prestasi akademik mahasiswa [11].

Penelitian berjudul “Perancangan Aplikasi K-Means Untuk Pengelompokan Mahasiswa STMIK Elrahma Yogyakarta Berdasarkan Frekuensi Kunjungan Ke Perpustakaan dan IPK” membahas tentang pengelompokan siswa didasarkan pada IPK dan frekuensi kunjungan ke perpustakaan dalam satu minggu. Penelitian ini juga menggunakan algoritma K-Means untuk menyelesaikan permasalahan yang ada. Hasil dari penelitian ini adalah diperoleh kesimpulan bahwa siswa sering pergi ke perpustakaan akan memiliki IPK (Indeks Prestasi rata-rata) cukup tinggi. Kelebihan penelitian ini yaitu dengan metode *clustering* K-Means dapat menyelesaikan masalah tentang pengelompokan mahasiswa. Kekurangan penelitian ini adalah hasil penelitian hanya menyimpulkan *cluster* berdasarkan 3 nilai IPK yaitu 1.48, 3.35, 3.49, sedangkan untuk *range*/kisaran IPK selain itu tidak masuk ke dalam kriteria sehingga tidak memiliki *cluster*. Alasan penggunaan metode *clustering* K-Means pada penelitian ini karena efektif dalam menyelesaikan masalah pengelompokan [12].

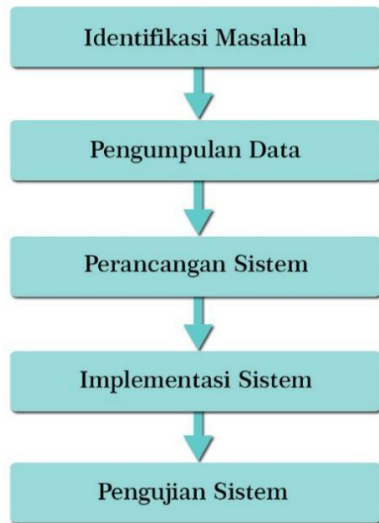
Penelitian berjudul “Sistem Informasi Pemetaan Siswa Berprestasi Di Bidang Olahraga Kecamatan Samarinda Seberang” membahas tentang sistem informasi pemetaan siswa berprestasi di bidang olahraga di Kecamatan Samarinda Seberang. Hasil dari penelitian ini adalah diambil kesimpulan bahwa telah dihasilkan sebuah sistem informasi pemetaan siswa sekolah berprestasi di bidang olahraga Kecamatan Samarinda Seberang. Kekurangan pada penelitian ini adalah masih berbasis desktop sehingga informasi yang ada terbatas untuk kalangan internal saja. Kelebihan penelitian ini adalah dapat menyelesaikan masalah yang ada yaitu pemetaan siswa berprestasi di Kecamatan Samarinda Seberang. Alasan penulis menggunakan Sistem Informasi Pemetaan yaitu karena ingin memberikan gambaran secara menyeluruh tentang letak sekolah, data sekolah, dan siswa berprestasi di setiap sekolah sehingga dapat memudahkan bagi pihak kecamatan saat memerlukan data pendidikan dan siswa berprestasi di Kecamatan Samarinda Seberang [13].

Penelitian berjudul “Sistem Informasi Geografis (SIG) Pemetaan Sekolah Berbasis Web Di Kecamatan Wonodadi Kabupaten Blitar” membahas tentang pemetaan sekolah di Kecamatan Wonodadi yang berbasis Web karena Kecamatan Wonodadi merupakan wilayah yang sedang berkembang dan memiliki penyediaan sarana fisik pendidikan yang memadai serta tenaga pendidik berkualitas. Penelitian ini menggunakan Sistem Informasi Geografis dengan bantuan *Google Map API* kemudian diimplementasikan ke *web*. Hasil dari penelitian ini adalah Sistem Informasi Geografis berbasis *web* yang menampilkan letak sekolah dari tingkat dasar, menengah pertama dan menengah atas. Kelebihan dari penelitian ini adalah Sistem Informasi Geografis berbasis *web* sehingga dapat diakses secara tepat dan efektif. Kekurangan penelitian ini yaitu belum berbasis Android karena lebih efisien dan lebih mudah diakses bagi penggunaannya. Sistem Informasi Geografis berbasis *web* dipilih karena dapat dengan mudah diakses bagi masyarakat di kecamatan Wonodadi dalam pencarian sekolah dan informasi dari sekolah tersebut [14].

Berdasarkan penelitian yang pernah dilakukan terkait dengan sistem informasi geografis dan algoritma K-Means, maka akan dilakukan penelitian yang membahas tentang Pemetaan Penyebaran Guru di Provinsi Banten. Adapun perbedaan antara penelitian terdahulu dengan penelitian yang dilakukan sekarang adalah bahwa dengan metode *clustering* K-Means akan dilakukan pemetaan penyebaran guru berdasarkan data jumlah guru, sekolah dan murid di Provinsi Banten pada tahun 2014-2015, kemudian dilakukan perhitungan dengan algoritma K-Means untuk dikelompokkan berdasarkan 3 *cluster* yaitu C1 (*Cluster 1*) dengan nilai kelebihan guru, C2 (*Cluster 2*) dengan nilai kecukupan guru dan C3 (*Cluster 3*) dengan nilai kekurangan guru sehingga pemerintah dapat mengetahui penyebaran guru untuk kemudian dilakukan pemerataan guru. Hasil dari pengelompokan tersebut akan dipetakan dengan menggunakan metode Sistem Informasi Geografis. Penelitian ini diharapkan dapat memberikan solusi dan saran dalam pemerataan penyebaran guru di tingkat SD, SMP dan SMA di Provinsi Banten.

III. METODE PENELITIAN

Secara umum penelitian terbagi ke dalam lima tahap, yaitu: (1) tahap identifikasi masalah, (2) tahap pengumpulan data, (3) tahap perancangan sistem, (4) tahap implementasi sistem, (5) tahap pengujian sistem.



Gambar 1. Tahapan Penelitian

Tahapan penelitian pada Gambar 1. dijelaskan sebagai berikut: Tahap identifikasi masalah: pada tahapan ini dilakukan analisis terhadap permasalahan yang ada, yaitu menyangkut masalah pemerataan guru pada tingkat SD, SMP dan SMA di provinsi Banten. Tahap pengumpulan data: pada tahapan ini pengumpulan data akan dilakukan dengan observasi. Tahap perancangan sistem: pada tahapan ini perancangan proses dalam sistem dilakukan dengan menggunakan algoritma K-Means. Tahap implementasi sistem: pada tahapan ini daerah di provinsi Banten dianalisis sesuai dengan perancangan sistem. Tahap pengujian sistem: Pada tahapan ini dilakukan pengujian sistem dengan menguji pengaruh dari *clustering*.

3.1. Tahap Identifikasi Masalah

Tahap pertama yang dilakukan adalah identifikasi masalah yaitu terkait masalah ketidakmerataan guru pada tingkat SD, SMP dan SMA di Provinsi Banten karena sebagian besar guru hanya menempati wilayah perkotaan. Berdasarkan data Badan Pusat Statistik tahun 2013, Provinsi Banten memiliki Indeks Pembangunan Manusia terendah di pulau Jawa. Guru memiliki peran penting dalam meningkatkan pendidikan karena guru merupakan pelaksana terdepan pendidikan anak-anak di sekolah. Dalam mengatasi ketidakmerataan guru maka pada penelitian ini dilakukan yaitu tentang pemetaan penyebaran guru pada tingkat SD, SMP dan SMA di Provinsi Banten.

3.1.1 Tahap Pengumpulan Data

Kemudian tahap selanjutnya adalah pengumpulan data tentang jumlah guru, jumlah murid, jumlah sekolah di Provinsi Banten dan peta Provinsi Banten.

3.1.2 Tahap Perancangan Sistem

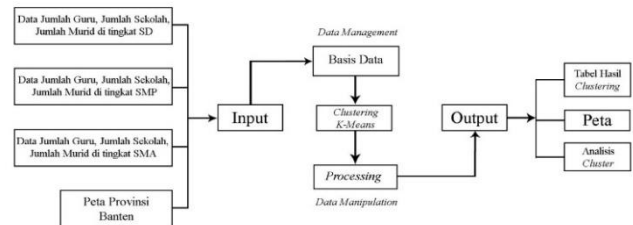
Tahap perancangan sistem menggunakan desain subsistem Sistem Informasi Geografis yang terdiri dari Data Input, Data Output, Data Management dan Data Manipulation.

3.1.3 Tahap Implementasi Sistem

Tahap implementasi sistem yaitu mengimplementasikan hasil *clustering* K-Means pada peta Provinsi Banten.

3.1.4 Tahap Pengujian Sistem

Tahap terakhir yaitu pengujian sistem dengan menggunakan analisis *cluster* untuk melihat faktor apa yang mempengaruhi pembentukan *cluster* yang ada.



Gambar 2. Diagram Desain Subsistem SIG

Dalam Sistem Informasi Geografis dapat diuraikan menjadi beberapa subsistem sebagai berikut:

- Data Input**
Data input pada sistem adalah: data jumlah guru, jumlah sekolah, jumlah murid pada tingkat SD, SMP dan SMA di Provinsi Banten serta peta Provinsi Banten.
- Data Output**
Subsistem output atau luaran sistem untuk menyajikan informasi seperti peta Provinsi Banten, tabel hasil *clustering* K-Means dan analisis *cluster*.
- Data Management**
Data *management* pada sistem bertugas untuk mengorganisasikan baik data spasial maupun tabel ke dalam sebuah basis data sehingga data dapat dengan mudah dikelompokkan menggunakan metode K-Means.
- Data Manipulation**
Data *manipulation* pada sistem bertugas dalam menentukan informasi-informasi yang dapat dihasilkan oleh Sistem Informasi Geografis dan pemodelan untuk menghasilkan informasi yang diharapkan.

IV. HASIL DAN PEMBAHASAN

Pseudocode adalah *Pseudocode* adalah kode atau tanda yang menyerupai (*pseudo*) atau penjelasan cara menyelesaikan suatu masalah. *Pseudocode* sering digunakan untuk menuliskan algoritma dari suatu permasalahan. *Pseudocode* berisikan langkah-langkah untuk menyelesaikan suatu permasalahan dan hampir sama dengan algoritma, hanya saja bentuknya sedikit berbeda dari algoritma. *Pseudocode* menggunakan bahasa yang hampir menyerupai bahasa pemrograman. Selain itu biasanya *Pseudocode* menggunakan bahasa yang mudah dipahami secara universal dan juga lebih ringkas dari pada algoritma.

Dalam kasus ini yaitu pemetaan guru di Provinsi Banten, data jumlah guru, jumlah murid dan jumlah sekolah dikelompokkan menggunakan metode *clustering* K-Means. *Pseudocode* pemetaan guru di Provinsi Banten dengan metode *clustering* K-Means dapat dilihat pada Kode Program 1.

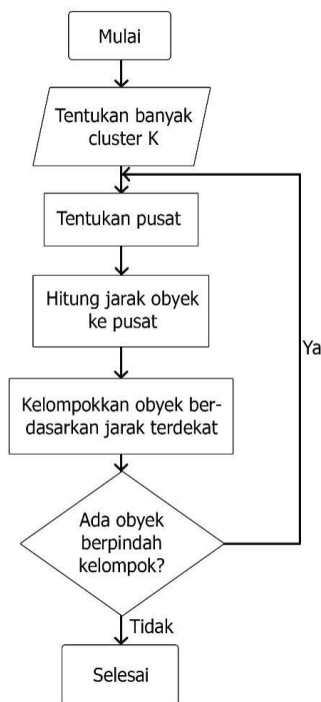
Kode Program 1. Pseudocode Clustering Data Guru di Provinsi Banten

```

1: Set  $ic$  (iteration count) to 1.
2: Determine total of clusters  $\{C_1, C_2, C_3, \dots, C_k\}$ .
3: Select randomly a set of  $K$  points  $m_1(1), m_2(1), m_3(1), \dots, m_k(1)$  as the initial centroids.
4: repeat
5:   For each vector  $x_i$  compute  $D(x_i, m_k(ic))$ , for each  $k = 1, \dots, k$  and assign  $x_i$  to the cluster  $C_j$  with the nearest mean.
6:   Increment  $ic$  by 1 and update the means to get new set  $m_1(ic), m_2(ic), m_3(1), \dots, m_k(ic)$ .
7: until  $C_k(ic) = C_k(ic+1)$  for all  $k$ .

```

K-Means termasuk dalam *partitioning clustering* yaitu setiap data harus masuk dalam *cluster* tertentu dan memungkinkan bagi setiap data yang termasuk dalam *cluster* tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke *cluster* yang lain.

**Gambar 3.** Tahap Algoritma K-Means

Langkah-langkah dalam melakukan algoritma K-Means:

1. Menentukan banyak *cluster*
Menentukan banyak *cluster* sesuai dengan masalah yang diteliti untuk mempermudah mengelompokkan data yang memiliki karakter sejenis.
2. Menentukan titik pusat
Selanjutnya menentukan titik pusat (*centroid*) dari tiap-tiap *cluster*. Pengambilan titik pusat secara *random* (acak).
3. Menghitung jarak setiap objek ke titik pusat

Kemudian menghitung jarak setiap data dengan titik pusat yang sudah ditentukan sebelumnya. Rumus untuk menghitung jarak setiap objek ke titik pusat adalah:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \quad (1)$$

4. Mengelompokkan objek berdasarkan jarak terdekat
Mengelompokkan objek berdasarkan jarak terdekat / minimum antara data dengan titik pusat.
5. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan sebagai berikut:

$$\mu = \frac{\sum_{i=1}^n x_i}{n}; i = 1, 2, 3, \dots, n \quad (2)$$

Dimana:

μ = titik pusat / *centroid* pada *cluster*

x_i = objek ke- i

n = banyaknya objek / jumlah objek yang menjadi anggota *cluster*

Jika pusat *cluster* tidak berubah lagi maka proses clustering selesai, atau kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.

Langkah pertama adalah menentukan jumlah *cluster* yang akan digunakan. Banyak *cluster* yang digunakan adalah 3 terdiri dari C1 (*Cluster* 1) dengan nilai kelebihan guru, C2 (*Cluster* 2) dengan nilai kecukupan guru dan C3 (*Cluster* 3) dengan nilai kekurangan guru. Data jumlah sekolah, murid dan guru di Provinsi Banten sesuai jenjang pendidikan dapat dilihat pada **Tabel 1.** pada tingkat SD, **Tabel 2.** pada tingkat SMP dan **Tabel 3.** pada tingkat SMA.

Tabel 1.

Jumlah Sekolah, Murid Dan Guru SD Di Provinsi Banten

No	Kabupaten/Kota	Sekolah	Murid	Guru
1.	Pandeglang	870	151231	9578
2.	Lebak	783	149897	7495
3.	Tangerang	943	314319	11225
4.	Serang	727	165162	7059
Kota				
5.	Tangerang	474	177606	6939
6.	Cilegon	175	45456	2386
7.	Serang	245	80664	2902
8.	Tangerang Selatan	342	130003	4752

Tabel 2.

Jumlah Sekolah, Murid Dan Guru SMP Di Provinsi Banten

No	Kabupaten/Kota	Sekolah	Murid	Guru
1.	Pandeglang	140	47193	1925
2.	Lebak	196	51699	1920
3.	Tangerang	330	132902	3158
4.	Serang	183	57012	1834
Kota				
5.	Tangerang	176	68544	2438
6.	Cilegon	41	14817	769
7.	Serang	71	27458	1022
8.	Tangerang Selatan	175	41919	1708

Tabel 3.

Jumlah Sekolah, Murid Dan Guru SMA Di Provinsi Banten

No	Kabupaten/Kota	Sekolah	Murid	Guru
1.	Pandeglang	34	18804	633
2.	Lebak	53	18438	670
3.	Tangerang	124	19341	1405
4.	Serang	67	29113	831
Kota				
5.	Tangerang	88	26981	1316
6.	Cilegon	20	6574	475
7.	Serang	30	9360	426
8.	Tangerang Selatan	65	20874	934

Langkah selanjutnya melakukan *clustering* dengan 3 *cluster* menggunakan Kode Program 4.2 untuk tingkat SD, Kode Program 4.3 untuk tingkat SMP dan Kode Program 4.4 untuk tingkat SMA.

Kode Program 2 Kode untuk *clustering* K-Means dengan 3 cluster (SD)

```
(SD <- kmeans(dataSD, 3))
```

```
> (SD <- kmeans(dataSD, 3))
K-means clustering with 3 clusters of sizes 1, 3, 5

Cluster means:
      GURU_SD MURID_SD SKLH_SD
1 11225.000 314319.0   943.0
2 1762.667  42040.0   140.0
3 7164.600 154779.8   639.2

Clustering vector:
[1] 3 3 2 3 3 1 3 3 2

within cluster sum of squares by cluster:
[1] 0 3275669501 1291164643
(between_SS / total_SS = 92.9 %)

Available components:
[1] "cluster" "centers" "totss" "withinss" "tot.withinss"
[6] "betweenss" "size" "iter" "ifault" "
```

Gambar 4. Hasil *clustering* K-Means di tingkat SD

Dapat dilihat hasil dari *clustering* K-Means pada **Gambar 4.** yang sudah dilakukan yaitu didapatkan "*cluster means*" yaitu nilai tengah pada data kelompok 'dataSD' yang dibagi menjadi 3 titik tengah/*centroid* karena data akan dikelompokkan menjadi 3 *cluster*. Kemudian data SD dihitung berdasarkan 3 titik tengah yang ada berdasarkan selisih yang paling sedikit. Lalu diperoleh "*clustering vector*" yaitu pengelompokkan data berdasarkan jarak titik tengah yang paling sedikit.

Kode Program 3. Kode untuk *clustering* K-Means dengan 3 cluster (SMP).

```
(SMP <- kmeans(dataSMP, 3))
```

```
> (SMP <- kmeans(dataSMP, 3))
K-means clustering with 3 clusters of sizes 3, 1, 5

Cluster means:
      GURU_SMP MURID_SMP SKLH_SMP
1 597 14091.67 37.33333
2 3158 132902.00 330.00000
3 1965 53273.40 174.00000

Clustering vector:
[1] 3 3 1 3 3 2 3 3 1

within cluster sum of squares by cluster:
[1] 378329203 0 415853043
(between_SS / total_SS = 93.1 %)

Available components:
[1] "cluster" "centers" "totss" "withinss" "tot.withinss"
[6] "betweenss" "size" "iter" "ifault" "
```

Gambar 5. Hasil *clustering* K-Means di tingkat SMP

Hasil dari *clustering* K-Means dapat dilihat pada **Gambar 5.** setelah dilakukan *clustering* yaitu didapatkan "*cluster means*" yaitu nilai tengah pada data kelompok 'dataSMP' yang dibagi menjadi 3 titik tengah / *centroid*. Kemudian data SMP dihitung berdasarkan 3 titik tengah yang ada berdasarkan selisih yang paling sedikit. Lalu diperoleh "*clustering vector*" yaitu pengelompokkan data berdasarkan jarak titik tengah yang paling sedikit.

Kode Program 4. Kode untuk *clustering* K-Means dengan 3 cluster (SMA).

```
(SMA <- kmeans(dataSMA, 3))
```



```
> (SMA <- kmeans(dataSMA,3))
K-means clustering with 3 clusters of sizes 4, 2, 3

cluster means:
  GURU_SMA MURID_SMA SKLH_SMA
1  910.5000 19364.250 69.00000
2 1073.5000 28047.000 77.50000
3  300.3333  5311.333 16.66667

clustering vector:
[1] 2 1 3 1 3 1 1 2 3

within cluster sum of squares by cluster:
[1] 3836156 2390545 46333258
(between_SS / total_SS = 92.8 %)

Available components:
[1] "cluster"      "centers"      "totss"        "withinss"
[5] "tot.withinss" "betweenss"    "size"         "iter"
[9] "ifault"
```

Gambar 6. Hasil *clustering* K-Means di tingkat SMA

Dapat dilihat hasil dari *clustering* K-Means pada **Gambar 6.** yang sudah dilakukan yaitu didapatkan “*cluster means*” yaitu nilai tengah pada data kelompok ‘dataSMA’ yang dibagi menjadi 3 titik tengah / *centroid*. Kemudian data SMA dihitung berdasarkan 3 titik tengah yang ada berdasarkan selisih yang paling sedikit. Lalu diperoleh “*clustering vector*” yaitu pengelompokan data berdasarkan jarak titik tengah yang paling sedikit. Kemudian menggabungkan data jumlah guru, jumlah murid, jumlah sekolah dengan *cluster*.

Kode Program 5. Kode untuk menggabungkan data dengan hasil *clustering* (SD).

```
dataSD$cluster <- SD$cluster
```

Kode Program 5. berfungsi untuk menggabungkan data jumlah guru SD, jumlah murid SD dan jumlah sekolah SD yang ada pada variabel dataSD dengan *cluster* yang telah didapat pada variabel SD. Hasil penggabungan data tersebut disimpan ke dalam variabel dataSD.

Kode Program 6. Kode untuk menggabungkan data dengan hasil *clustering* (SMP).

```
dataSMP$cluster <- SMP$cluster
```

Kode Program 6. berfungsi untuk menggabungkan data jumlah guru SMP, jumlah murid SMP dan jumlah sekolah SMP yang ada pada variabel dataSMP dengan *cluster* yang telah didapat pada variabel SMP. Hasil penggabungan data tersebut disimpan ke dalam variabel dataSMP.

Kode Program 7. Kode untuk menggabungkan data dengan hasil *clustering* (SMA).

```
dataSMA$cluster <- SMA$cluster
```

Kode Program 7. berfungsi untuk menggabungkan data jumlah guru SMA, jumlah murid SMA dan jumlah sekolah SMA yang ada pada variabel dataSD dengan *cluster* yang telah didapat pada variabel SMA. Hasil penggabungan data tersebut disimpan ke dalam variabel dataSMA. Langkah selanjutnya mengubah data peta menjadi *dataframe* agar dapat diplot ke dalam peta.

Kode Program 8. Mengubah *shapefile* peta SD ke dalam *dataframe*.

```
states.SD.f <- fortify(states.SD, region = 'CLUSTER')
```

Kode Program 8. berfungsi mengubah data peta SD menjadi *dataframe* berdasarkan *region* / kabupaten yang terdapat dalam peta SD agar hasil pengelompokan menggunakan *clustering* K-Means dapat diplot menjadi peta.

Kode Program 9. Mengubah *shapefile* peta SMP ke dalam *dataframe*

```
states.SMP.f <- fortify(states.SMP, region = 'CLUSTER')
```

Kode Program 9. berfungsi mengubah data peta SMP menjadi *dataframe* berdasarkan *region* / kabupaten yang terdapat dalam peta SMP agar hasil pengelompokan menggunakan *clustering* K-Means dapat diplot menjadi peta.

Kode Program 10. Mengubah *shapefile* peta SMA ke dalam *dataframe*

```
states.SMA.f <- fortify(states.SMA, region = 'CLUSTER')
```

Kode Program 10. berfungsi mengubah data peta SMA menjadi *dataframe* berdasarkan *region* / kabupaten yang terdapat dalam peta SMA agar hasil pengelompokan menggunakan *clustering* K-Means dapat diplot menjadi peta. Kemudian menggabungkan data peta dalam bentuk *dataframe* dengan data yang telah diolah dengan *clustering* K-Means.

Kode Program 11. Kode untuk menggabungkan dataSD dan data peta.

```
merge.SD.coef<-merge(states.SD.f, dataSD, by="id", all.x=TRUE)
```

Menggabungkan peta Provinsi Banten di tingkat SD yang telah diubah menjadi *dataframe* pada variabel states.SD.f dengan dataSD yang berisi jumlah guru, jumlah murid, jumlah sekolah dan *cluster* berdasarkan id lalu disimpan ke dalam variabel merge.SD.coef seperti pada Kode Program 12.

Kode Program 12. Kode untuk menggabungkan dataSMP dan data peta

```
merge.SMP.coef<-merge(states.SMP.f, dataSMP, by="id", all.x=TRUE)
```

Menggabungkan peta Provinsi Banten di tingkat SMP yang telah diubah menjadi *dataframe* pada variabel states.SMP.f dengan dataSMP yang berisi jumlah guru, jumlah murid, jumlah sekolah dan *cluster* berdasarkan id lalu

disimpan ke dalam variabel `merge.SMP.coef` seperti pada Kode Program 12.

Kode Program 13. Kode untuk menggabungkan data SMA dan data peta

```
merge.SMA.coef<-merge(states.SMA.f,
dataSMA, by="id", all.x=TRUE)
```

Menggabungkan peta Provinsi Banten di tingkat SMA yang telah diubah menjadi *dataframe* pada variabel `states.SD.f` dengan data SMA yang berisi jumlah guru, jumlah murid, jumlah sekolah dan *cluster* berdasarkan id lalu disimpan ke dalam variabel `merge.SMA.coef` seperti pada Kode Program 13. Lalu mengurutkan data dengan menggunakan perintah *order* berdasarkan urutan dari variabel *order* yaitu urutan berdasarkan id terkecil.

Kode Program 14. Kode untuk mengurutkan data di tingkat SD.

```
final.plotSD<-
merge.SD.coef[order(merge.SD.coef$order), ]
```

Mengurutkan data di tingkat SD pada variabel `merge.SD.coef` berdasarkan variabel *order* / urutan nilai yang paling kecil menggunakan perintah *order* dan disimpan pada variabel `final.plotSD` seperti pada Kode Program 14.

Kode Program 15. Kode untuk mengurutkan data di tingkat SMP.

```
final.plotSMP<-
merge.SMP.coef[order(merge.SMP.coef$order), ]
```

Mengurutkan data di tingkat SMP pada variabel `merge.SMP.coef` berdasarkan variabel *order* / urutan nilai yang paling kecil menggunakan perintah *order* dan disimpan pada variabel `final.plotSMP` seperti pada Kode Program 15.

Kode Program 16. Kode untuk mengurutkan data di tingkat SMA.

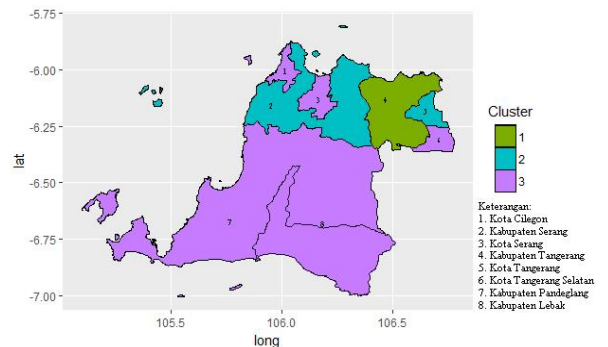
```
final.plotSMA<-
merge.SMA.coef[order(merge.SMA.coef$order), ]
```

Mengurutkan data di tingkat SMA pada variabel `merge.SMA.coef` berdasarkan variabel *order* / urutan nilai yang paling kecil menggunakan perintah *order* dan disimpan pada variabel `final.plotSMA` seperti pada Kode Program 16.

Langkah terakhir yaitu menampilkan peta sesuai dengan jenjang pendidikan sesuai dengan *cluster* yang ada. Untuk menampilkan peta pada tingkat SD dapat dilihat pada Kode Program 17. untuk tingkat SMP dapat dilihat pada Kode Program 18. dan untuk tingkat SMA dapat dilihat pada Kode Program 19.

Kode Program 17. *Pseudocode* untuk menampilkan peta di tingkat SD.

```
1: Set data using final.plotSD
2: Set long as x axis
3: Set lat as y axis
4: Set black color as outline of each district
5: Fill map according to data cluster
6: Plot map
```

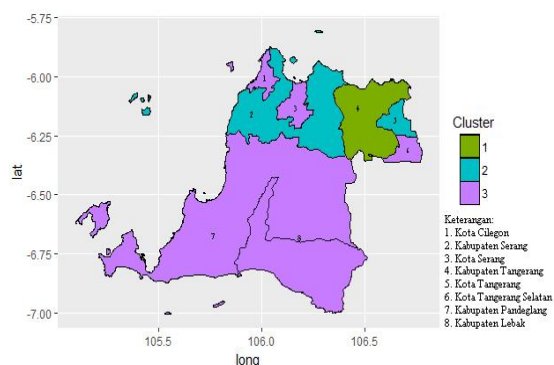


Gambar 7. Hasil *plotting* peta di tingkat SD

Hasil plot peta di tingkat SD dapat dilihat pada Gambar 7. daerah yang berwarna hijau adalah daerah yang masuk pada *cluster* 1 yaitu kelebihan guru yaitu seperti daerah Kabupaten Tangerang. Daerah yang berwarna biru masuk ke dalam *cluster* 2 yaitu kecukupan guru seperti daerah Kabupaten Serang dan Kota Tangerang. Lalu daerah yang berwarna ungu adalah daerah yang masuk pada *cluster* 3 dengan kekurangan guru seperti daerah Kota Cilegon, Kota Serang, Kota Tangerang Selatan, Kabupaten Pandeglang dan Kabupaten Lebak.

Kode Program 18. *Pseudocode* untuk menampilkan peta di tingkat SMP

```
1: Set data using final.plotSMP
2: Set long as x axis
3: Set lat as y axis
4: Set black color as outline of each district
5: Fill map according to data cluster
6: Plot map
```

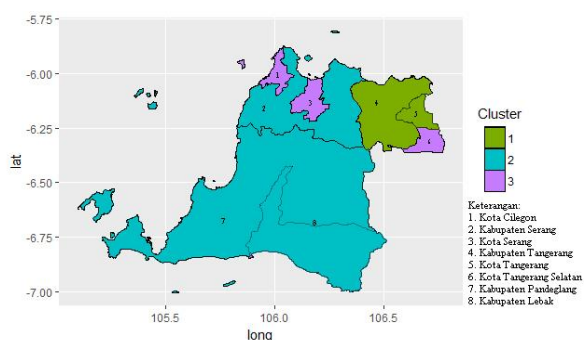


Gambar 8. Hasil plotting peta di tingkat SMP

Hasil plot peta di tingkat SMP dapat dilihat pada Gambar 8. daerah yang berwarna hijau adalah daerah yang masuk pada *cluster 1* yaitu kelebihan guru yaitu seperti daerah Kabupaten Tangerang. Daerah yang berwarna biru masuk ke dalam *cluster 2* yaitu kecukupan guru seperti daerah Kabupaten Serang dan Kota Tangerang. Lalu daerah yang berwarna ungu adalah daerah yang masuk pada *cluster 3* dengan kekurangan guru seperti daerah Kota Cilegon, Kota Serang, Kota Tangerang Selatan, Kabupaten Pandeglang dan Kabupaten Lebak.

Kode Program 9. Pseudocode untuk menampilkan peta di tingkat SMA

```
1: Set data using final.plotSMA
2: Set long as x axis
3: Set lat as y axis
4: Set black color as outline of each district
5: Fill map according to data cluster
6: Plot map
```



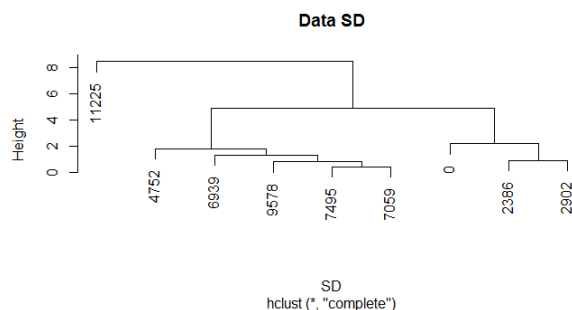
Gambar 9. Hasil plotting peta di tingkat SMA

Hasil plot peta di tingkat SMA dapat dilihat pada Gambar 9. daerah yang berwarna hijau adalah daerah yang masuk pada *cluster 1* yaitu kelebihan guru yaitu seperti daerah Kabupaten Tangerang dan Kota Tangerang. Daerah yang berwarna biru masuk ke dalam *cluster 2* yaitu kecukupan guru seperti daerah Kabupaten Serang, Kabupaten Pandeglang dan Kabupaten Lebak. Lalu daerah yang berwarna ungu adalah daerah yang masuk pada *cluster 3*

dengan kekurangan guru seperti daerah Kota Cilegon, Kota Serang, Kota Tangerang Selatan.

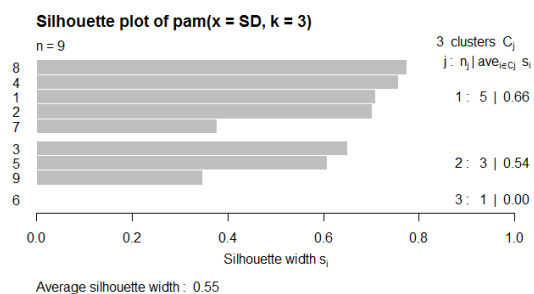
4.1 Analisis Cluster K-Means

1. Data SD



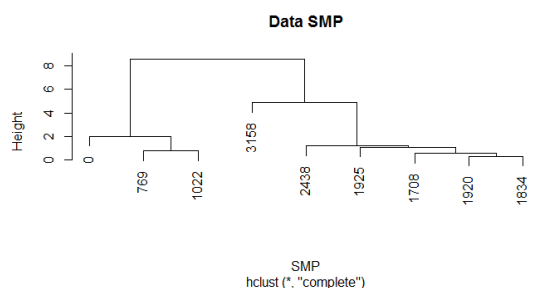
Gambar 10. Pengelompokan data guru SD

Berdasarkan hasil yang diperoleh pada Gambar 10. dapat ditarik kesimpulan bahwa data untuk tingkat SD yang memiliki selisih jarak paling sedikit akan masuk ke dalam 1 *cluster* yang sama, dengan kata lain data yang memiliki karakteristik sejenis akan masuk ke dalam *cluster* yang sama. Kemudian pada Gambar 11. dapat dilihat nilai rata-rata dari setiap *cluster* yang ada.



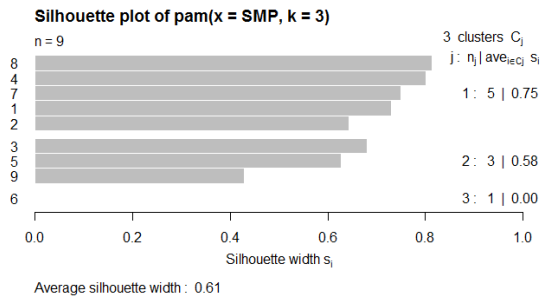
Gambar 11. Hasil plotting pengelompokan guru SD berdasarkan *cluster* sejenis

2. Tingkat SMP



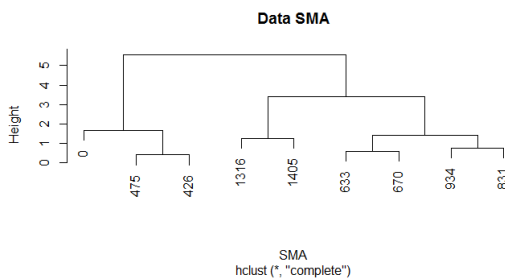
Gambar 12. Pengelompokan data guru SMP

Berdasarkan hasil yang diperoleh pada Gambar 12. dapat ditarik kesimpulan bahwa data untuk tingkat SD yang memiliki selisih jarak paling sedikit akan masuk ke dalam 1 *cluster* yang sama, dengan kata lain data yang memiliki karakteristik sejenis akan masuk ke dalam *cluster* yang sama. Kemudian pada Gambar 15. dapat dilihat nilai rata-rata dari setiap *cluster* yang ada.



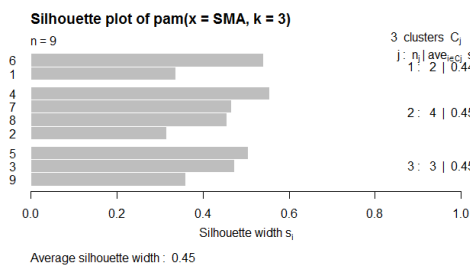
Gambar 13. Hasil *plotting* pengelompokkan guru SMP berdasarkan *cluster* sejenis

3. Tingkat SMA

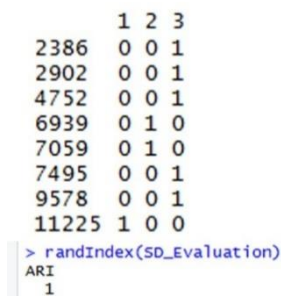


Gambar 14. Pengelompokkan data guru SMA

Berdasarkan hasil yang diperoleh pada Gambar 14. dapat ditarik kesimpulan bahwa data untuk tingkat SD yang memiliki selisih jarak paling sedikit akan masuk ke dalam 1 *cluster* yang sama, dengan kata lain data yang memiliki karakteristik sejenis akan masuk ke dalam *cluster* yang sama. Kemudian pada Gambar 15. dapat dilihat nilai rata-rata dari setiap *cluster* yang ada.

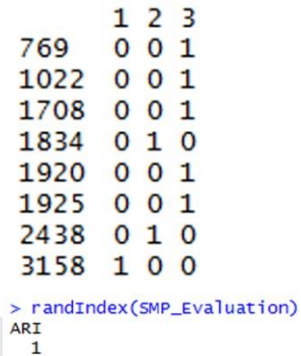


Gambar 15. Hasil *plotting* pengelompokkan guru SMA berdasarkan *cluster* sejenis



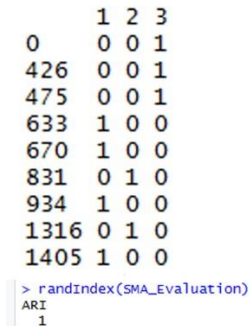
Gambar 16. Evaluasi hasil *clustering* di tingkat SD

Berdasarkan *clustering* yang telah dilakukan pada tingkat SD diperoleh kesimpulan bahwa hasil *clustering* yang diperoleh memiliki ketepatan baik yaitu sebesar 1 atau 100% dapat dilihat pada Gambar 16.



Gambar 17. Evaluasi hasil *clustering* di tingkat SMP

Berdasarkan *clustering* yang telah dilakukan pada tingkat SMP diperoleh kesimpulan bahwa hasil *clustering* yang diperoleh memiliki ketepatan baik yaitu sebesar 1 atau 100% dapat dilihat pada Gambar 17.



Gambar 18. Evaluasi hasil *clustering* di tingkat SMA

Berdasarkan *clustering* yang telah dilakukan pada tingkat SMA diperoleh kesimpulan bahwa hasil *clustering* yang diperoleh memiliki ketepatan baik yaitu sebesar 1 atau 100% dapat dilihat pada Gambar 18.

V. SIMPULAN

Berdasarkan masalah yang ada yaitu mengenai masalah pemerataan guru di tingkat SD, SMP dan SMA di Provinsi Banten, hasil penelitian ini diharapkan dapat menyelesaikan masalah tersebut. Dari penelitian yang telah dilakukan, hasil yang diperoleh adalah sebagai berikut: (1) Hasil plot peta di tingkat SD menunjukkan bahwa Kabupaten Tangerang memiliki kelebihan guru. Kabupaten Serang dan Kota Tangerang memiliki kecukupan guru sedangkan daerah Kota Cilegon, Kota Serang, Kota Tangerang Selatan, Kabupaten Pandeglang dan Kabupaten Lebak memiliki kekurangan guru; (2) Hasil plot peta di tingkat SMP menunjukkan bahwa Kabupaten Tangerang memiliki kelebihan guru. Daerah Kabupaten Serang dan Kota Tangerang memiliki kecukupan guru dan daerah Kota Cilegon, Kota Serang, Kota Tangerang Selatan, Kabupaten Pandeglang dan Kabupaten Lebak memiliki kekurangan guru; (3) Hasil plot peta di tingkat SMA menunjukkan bahwa daerah Kabupaten Tangerang dan Kota Tangerang memiliki

kelebihan guru. Daerah Kabupaten Serang, Kabupaten Pandeglang dan Kabupaten Lebak memiliki kecukupan guru. Sedangkan daerah Kota Cilegon, Kota Serang, Kota Tangerang Selatan memiliki kekurangan guru.

Sedangkan hasil dari analisis *cluster* yang telah dilakukan menunjukkan bahwa atribut yang memiliki nilai sejenis atau jarak nilai satu ke nilai lain berdekatan maka akan dikelompokkan ke dalam 1 *cluster*. Pemetaan guru di Provinsi Banten diharapkan dapat menjadi rekomendasi Dinas Pendidikan Provinsi Banten dalam hal pemerataan guru baik di tingkat SD, SMP dan SMA. Saran dari penulis adalah jika ingin melakukan *clustering* data yang digunakan harus lebih banyak dan penelitian selanjutnya dapat dikembangkan dengan menggunakan algoritma *Fuzzy C-Means*.

DAFTAR PUSTAKA

- [1] Muhammad, 2015. *Kualitas Pendidikan Indonesia di Mata Dunia*. Diambil 01 Juli 2017 dari <https://www.taralite.com/artikel/post/kualitas-pendidikan-indonesia-di-mata-dunia/>.
- [2] Rusdi, A., 2014. *Distribusi Guru Tak Merata Penyebab Mutu Pendidikan Rendah*. Diambil 30 Juni 2017 dari <https://www.ajnn.net/news/distribusi-guru-tak-merata-penyebab-mutu-pendidikan-rendah/index.html>.
- [3] BPH UMY, 2012. *Anies Baswedan: Distribusi Guru di Indonesia Belum Merata*. Diambil 01 Juli 2017 dari <http://www.umi.ac.id/anies-baswedan-distribusi-guru-di-indonesia-belum-merata.html>.
- [4] Aparatur Sipil Negara, 2014. *Data Kelebihan dan Kekurangan Guru di Indonesia*. Diambil 30 Juni 2017 dari <http://www.asncpns.com/2014/04/data-kelebihan-dan-kekurangan-guru-di.html>.
- [5] BPS, 2013. *Indeks Pembangunan Manusia Menurut Provinsi, 1996–2013*. Diambil 30 Juni 2017 dari <https://www.bps.go.id/linkTableDinamis/view/id/909>.
- [6] SatelitNews, 2012. *60 Persen Guru PNS Banten Numpuk Diperkotaan*. Diambil 22 Juni 2017 <http://satelitnews.co.id/2012/03/16/60-persen-guru-pns-banten-numpuk-diperkotaan>.
- [7] Panjaitan, M. *Implementasi Algoritma K-Means Dan Analytic Hierarchy Process (AHP) Untuk Klasterisasi Guru Dan Memilih Guru Terbaik*. Diambil 22 Juni 2017 dari https://www.academia.edu/9684064/Implementasi_Algoritma_K-Means_Dan_Analytic_Hierarchy_Process_AHP_Untuk_Klasterisasi_Guru_Dan_Memilih_Guru_Terbaik.
- [8] Aronoff, S, 1989. *Mengenal SIG dan Data Spasial*. Diambil 24 Juni 2016 dari <http://osgeo.ft.ugm.ac.id/mengenal-sig-dan-data-spasial/>.
- [9] Samino, 2006, *Sistem Informasi Geografis Sekolah Menengah Atas (SMA/SMK) di Yogyakarta*, Skripsi S-1 Universitas Ahmad Dahlan.
- [10] Kharistiani, E, 2013, Juni. *Sistem Informasi Geografis Pemetaan Potensi SMA/SMK Berbasis Web*, *Jurnal Sarjana Teknik Informatika*, Vol. 01 No. 01.
- [11] Oyelade, O. J., Oladipupo, O. O., Obagbuwa, I. C., 2010. *Application of k-Means Clustering algorithm for prediction of Students' Academic Performance*, *International Journal of Computer Science and Information Security*, Vol. 07 No. 01.
- [12] Syafrianto, A, 2012. *Perancangan Aplikasi K-Means Untuk Pengelompokan Mahasiswa STMIK Elrahma Yogyakarta Berdasarkan Frekuensi Kunjungan Ke Perpustakaan Dan IPK*. Skripsi. Yogyakarta: Fakultas Teknik Informatika, STMIK EL-RAHMA Yogyakarta.
- [13] Ihsan, M. S., Dengen, N., Hairah, U., 2017. *Sistem Informasi Pemetaan Siswa Berprestasi Di Bidang Olahraga Kecamatan Samarinda Seberang*, *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, Vol. 02 No. 01.
- [14] Husaini, M. A., 2017. *Sistem Informasi Geografis (GIS) Pemetaan Sekolah Berbasis Web Di Kecamatan Wonodadi Kabupaten Blitar*, *Jurnal Antivirus*, Vol. 11 No. 01.
- [15] Santosa, B, 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta : Graha Ilmu.